WORLDWIDE PDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 24, 2017 – 07:29 AM EDT

PDB ID	:	3GG2
Title	:	Crystal structure of UDP-glucose 6-dehydrogenase from Porphyromonas gingivalis bound to product UDP-glucuronate
Authors	:	Bonanno, J.B.; Freeman, J.; Bain, K.T.; Chang, S.; Sampathkumar, P.; Wasserman, S.; Sauder, J.M.; Burley, S.K.; Almo, S.C.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on	:	unknown
Resolution	:	1.70 Å(reported)

The following versions of software and data (see references ⓘ) were used in the production of this report:

|		
MolProbity	:	4.02b-467
Mogul	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20030345

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.70 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Percentile Ranks	Value
Rfree		0.258
Clashscore		4
Ramachandran outliers		0
Sidechain outliers		2.4%
RSRZ outliers		3.9%

Worse   *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution
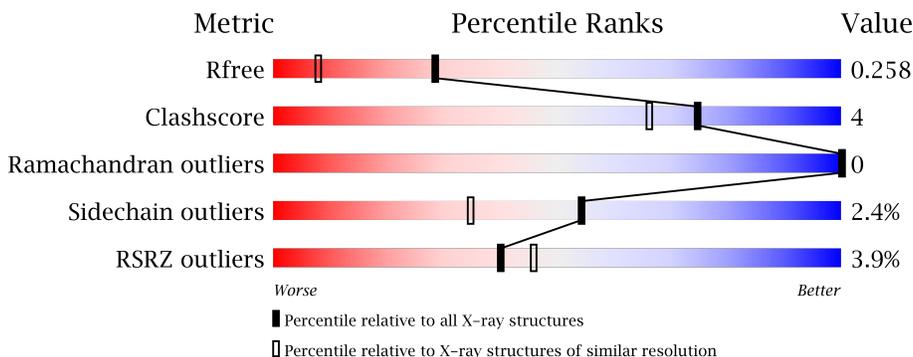
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	100719	3453 (1.70-1.70)
Clashscore	112137	3876 (1.70-1.70)
Ramachandran outliers	110173	3815 (1.70-1.70)
Sidechain outliers	110143	3815 (1.70-1.70)
RSRZ outliers	101464	3491 (1.70-1.70)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	450	3% 87% 8% · ·
1	B	450	6% 84% 9% · 6%
1	C	450	3% 85% 8% 7%
1	D	450	2% 88% 6% · 5%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 14679 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Sugar dehydrogenase, UDP-glucose/GDP-mannose dehydrogenase family.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	431	Total	C	N	O	S	0	8	0
			3360	2136	570	631	23			
1	B	425	Total	C	N	O	S	0	2	0
			3256	2065	551	617	23			
1	C	420	Total	C	N	O	S	0	3	0
			3229	2048	545	613	23			
1	D	427	Total	C	N	O	S	0	4	0
			3305	2101	561	620	23			

There are 60 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	77	MET	-	expression tag	UNP Q7MVC7
A	78	SER	-	expression tag	UNP Q7MVC7
A	79	LEU	-	expression tag	UNP Q7MVC7
A	158	ILE	VAL	engineered	UNP Q7MVC7
A	175	SER	GLY	engineered	UNP Q7MVC7
A	211	ALA	VAL	engineered	UNP Q7MVC7
A	485	ALA	THR	engineered	UNP Q7MVC7
A	519	GLU	-	expression tag	UNP Q7MVC7
A	520	GLY	-	expression tag	UNP Q7MVC7
A	521	HIS	-	expression tag	UNP Q7MVC7
A	522	HIS	-	expression tag	UNP Q7MVC7
A	523	HIS	-	expression tag	UNP Q7MVC7
A	524	HIS	-	expression tag	UNP Q7MVC7
A	525	HIS	-	expression tag	UNP Q7MVC7
A	526	HIS	-	expression tag	UNP Q7MVC7
B	77	MET	-	expression tag	UNP Q7MVC7
B	78	SER	-	expression tag	UNP Q7MVC7
B	79	LEU	-	expression tag	UNP Q7MVC7
B	158	ILE	VAL	engineered	UNP Q7MVC7
B	175	SER	GLY	engineered	UNP Q7MVC7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	211	ALA	VAL	engineered	UNP Q7MVC7
B	485	ALA	THR	engineered	UNP Q7MVC7
B	519	GLU	-	expression tag	UNP Q7MVC7
B	520	GLY	-	expression tag	UNP Q7MVC7
B	521	HIS	-	expression tag	UNP Q7MVC7
B	522	HIS	-	expression tag	UNP Q7MVC7
B	523	HIS	-	expression tag	UNP Q7MVC7
B	524	HIS	-	expression tag	UNP Q7MVC7
B	525	HIS	-	expression tag	UNP Q7MVC7
B	526	HIS	-	expression tag	UNP Q7MVC7
C	77	MET	-	expression tag	UNP Q7MVC7
C	78	SER	-	expression tag	UNP Q7MVC7
C	79	LEU	-	expression tag	UNP Q7MVC7
C	158	ILE	VAL	engineered	UNP Q7MVC7
C	175	SER	GLY	engineered	UNP Q7MVC7
C	211	ALA	VAL	engineered	UNP Q7MVC7
C	485	ALA	THR	engineered	UNP Q7MVC7
C	519	GLU	-	expression tag	UNP Q7MVC7
C	520	GLY	-	expression tag	UNP Q7MVC7
C	521	HIS	-	expression tag	UNP Q7MVC7
C	522	HIS	-	expression tag	UNP Q7MVC7
C	523	HIS	-	expression tag	UNP Q7MVC7
C	524	HIS	-	expression tag	UNP Q7MVC7
C	525	HIS	-	expression tag	UNP Q7MVC7
C	526	HIS	-	expression tag	UNP Q7MVC7
D	77	MET	-	expression tag	UNP Q7MVC7
D	78	SER	-	expression tag	UNP Q7MVC7
D	79	LEU	-	expression tag	UNP Q7MVC7
D	158	ILE	VAL	engineered	UNP Q7MVC7
D	175	SER	GLY	engineered	UNP Q7MVC7
D	211	ALA	VAL	engineered	UNP Q7MVC7
D	485	ALA	THR	engineered	UNP Q7MVC7
D	519	GLU	-	expression tag	UNP Q7MVC7
D	520	GLY	-	expression tag	UNP Q7MVC7
D	521	HIS	-	expression tag	UNP Q7MVC7
D	522	HIS	-	expression tag	UNP Q7MVC7
D	523	HIS	-	expression tag	UNP Q7MVC7
D	524	HIS	-	expression tag	UNP Q7MVC7
D	525	HIS	-	expression tag	UNP Q7MVC7
D	526	HIS	-	expression tag	UNP Q7MVC7

- Molecule 2 is URIDINE-5'-DIPHOSPHATE-GLUCURONIC ACID (three-letter code: UGA) (formula: $C_{15}H_{22}N_2O_{18}P_2$).

UGA
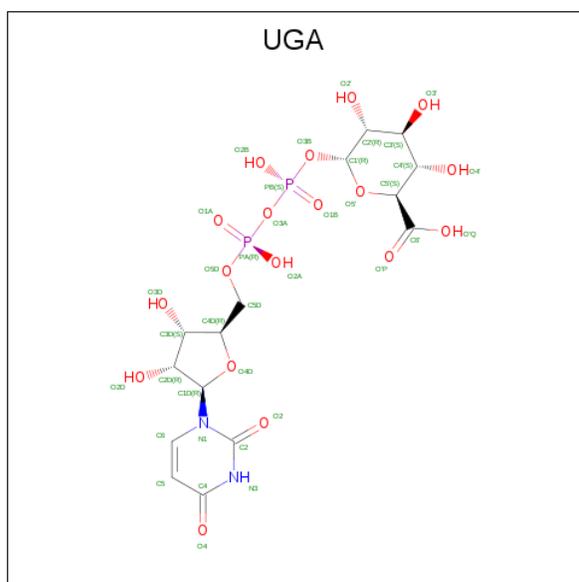
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	P	0	0
			37	15	2	18	2		
2	A	1	Total	C	N	O	P	0	0
			37	15	2	18	2		
2	B	1	Total	C	N	O	P	0	0
			37	15	2	18	2		
2	C	1	Total	C	N	O	P	0	0
			37	15	2	18	2		
2	D	1	Total	C	N	O	P	0	0
			37	15	2	18	2		
2	D	1	Total	C	N	O	P	0	0
			37	15	2	18	2		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	362	Total	O	0	0
			362	362		
3	B	295	Total	O	0	0
			295	295		
3	C	366	Total	O	0	0
			366	366		
3	D	284	Total	O	0	0
			284	284		

# 3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: Sugar dehydrogenase, UDP-glucose/GDP-mannose dehydrogenase family
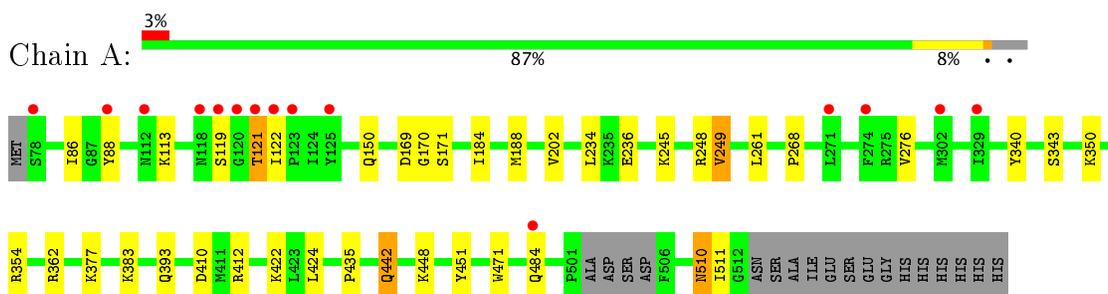
- Molecule 1: Sugar dehydrogenase, UDP-glucose/GDP-mannose dehydrogenase family
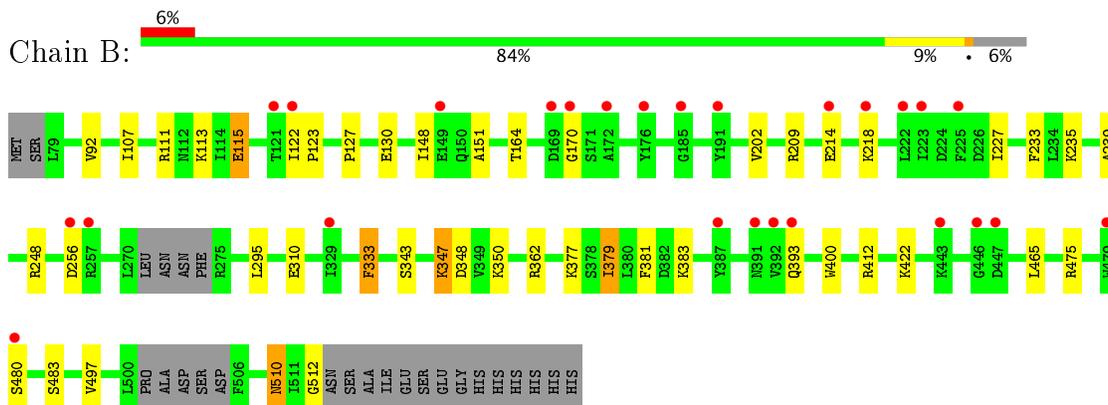

- Molecule 1: Sugar dehydrogenase, UDP-glucose/GDP-mannose dehydrogenase family
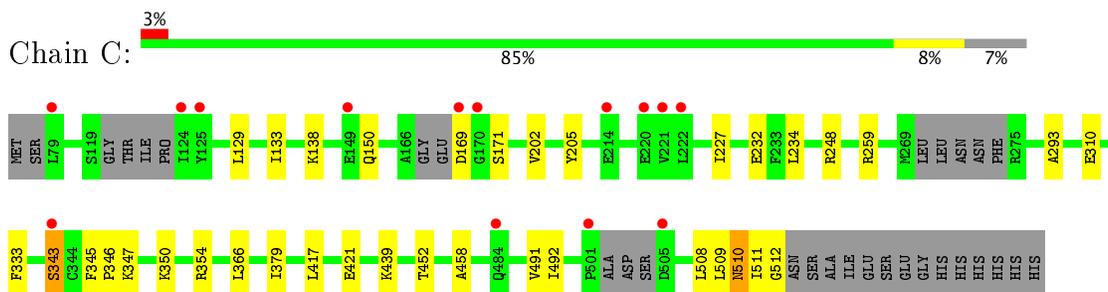
- Molecule 1: Sugar dehydrogenase, UDP-glucose/GDP-mannose dehydrogenase family
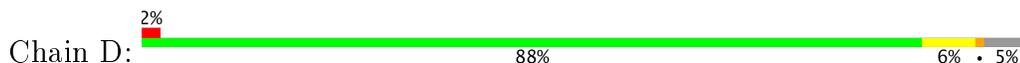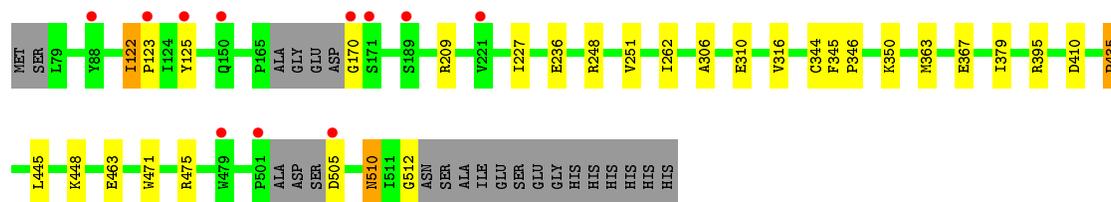
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	82.36Å 143.49Å 85.50Å 90.00° 112.46° 90.00°	Depositor
Resolution (Å)	20.00 – 1.70 44.35 – 1.70	Depositor EDS
% Data completeness (in resolution range)	93.7 (20.00-1.70) 93.3 (44.35-1.70)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	0.12	Depositor
$<I/\sigma(I)>$ [1]	2.76 (at 1.70Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.215 , 0.256 0.218 , 0.258	Depositor DCC
R_{free} test set	9459 reflections (5.32%)	DCC
Wilson B-factor (Å^2)	17.8	Xtriage
Anisotropy	0.134	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 53.4	EDS
L-test for twinning[2]	$<|L|> = 0.51$, $<L^2> = 0.34$	Xtriage
Estimated twinning fraction	0.022 for l,-k,h	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	14679	wwPDB-VP
Average B, all atoms (Å^2)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.75% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $<|L|>$, $<L^2>$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: UGA

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$|Z|$>5	RMSZ	#$|Z|$>5
1	A	0.75	0/3438	0.76	0/4642
1	B	0.67	0/3312	0.72	1/4476 (0.0%)
1	C	0.70	0/3286	0.74	0/4438
1	D	0.70	0/3370	0.73	1/4553 (0.0%)
All	All	0.71	0/13406	0.74	2/18109 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	344	CYS	CA-CB-SG	-5.43	104.23	114.00
1	B	377	LYS	CD-CE-NZ	5.01	123.21	111.70

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3360	0	3402	29	0
1	B	3256	0	3256	29	1
1	C	3229	0	3222	23	0
1	D	3305	0	3322	17	0
2	A	74	0	38	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	37	0	19	0	0
2	C	37	0	19	0	0
2	D	74	0	38	1	0
3	A	362	0	0	4	0
3	B	295	0	0	4	0
3	C	366	0	0	5	0
3	D	284	0	0	1	1
All	All	14679	0	13316	98	1

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 4.

The worst 5 of 98 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:448:LYS:HG3	3:A:1815:HOH:O	1.65	0.94
1:C:202:VAL:HG23	3:C:560:HOH:O	1.74	0.87
1:B:107:ILE:HD11	1:B:151:ALA:HB2	1.61	0.82
1:B:107:ILE:HD11	1:B:151:ALA:CB	2.10	0.81
1:B:233:PHE:O	3:B:2215:HOH:O	2.02	0.78

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:362:ARG:NH2	3:D:1429:HOH:O[1_655]	2.13	0.07

5.3   Torsion angles ⓘ

5.3.1   Protein backbone ⓘ

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	435/450 (97%)	430 (99%)	5 (1%)	0	100	100
1	B	421/450 (94%)	416 (99%)	5 (1%)	0	100	100
1	C	413/450 (92%)	406 (98%)	7 (2%)	0	100	100
1	D	425/450 (94%)	419 (99%)	6 (1%)	0	100	100
All	All	1694/1800 (94%)	1671 (99%)	23 (1%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	360/374 (96%)	348 (97%)	12 (3%)	43	21
1	B	343/374 (92%)	332 (97%)	11 (3%)	44	22
1	C	341/374 (91%)	335 (98%)	6 (2%)	64	47
1	D	351/374 (94%)	346 (99%)	5 (1%)	71	58
All	All	1395/1496 (93%)	1361 (98%)	34 (2%)	54	35

5 of 34 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	256	ASP
1	B	393	GLN
1	D	435	PRO
1	B	347	LYS
1	A	362	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	510	ASN
1	D	510	ASN
1	C	150	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	510	ASN
1	C	510	ASN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

6 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#|Z| > 2$	Counts	RMSZ	$\#|Z| > 2$
2	UGA	A	1000	-	32,39,39	2.15	3 (9%)	44,60,60	2.22	7 (15%)
2	UGA	A	6000	-	32,39,39	2.27	3 (9%)	44,60,60	1.96	5 (11%)
2	UGA	B	2000	-	32,39,39	2.08	3 (9%)	44,60,60	2.21	7 (15%)
2	UGA	C	3000	-	32,39,39	1.97	3 (9%)	44,60,60	2.34	9 (20%)
2	UGA	D	4000	-	32,39,39	2.31	5 (15%)	44,60,60	2.02	5 (11%)
2	UGA	D	5000	-	32,39,39	2.31	3 (9%)	44,60,60	1.94	4 (9%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns.

'-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	UGA	A	1000	-	-	0/21/61/61	0/3/3/3
2	UGA	A	6000	-	-	0/21/61/61	0/3/3/3
2	UGA	B	2000	-	-	0/21/61/61	0/3/3/3
2	UGA	C	3000	-	-	0/21/61/61	0/3/3/3
2	UGA	D	4000	-	-	0/21/61/61	0/3/3/3
2	UGA	D	5000	-	-	0/21/61/61	0/3/3/3

The worst 5 of 20 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	5000	UGA	C6-N1	-9.72	1.34	1.47
2	A	6000	UGA	C6-N1	-9.04	1.35	1.47
2	D	4000	UGA	C6-N1	-8.28	1.36	1.47
2	D	4000	UGA	C6-C5	-7.71	1.38	1.52
2	C	3000	UGA	C6-N1	-7.66	1.37	1.47

The worst 5 of 37 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	3000	UGA	C4-N3-C2	-7.54	119.35	125.81
2	A	1000	UGA	C4-N3-C2	-6.59	120.16	125.81
2	A	6000	UGA	C4-N3-C2	-6.18	120.51	125.81
2	B	2000	UGA	C4-N3-C2	-5.51	121.09	125.81
2	D	4000	UGA	C4-N3-C2	-5.25	121.31	125.81

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	6000	UGA	2	0
2	D	4000	UGA	1	0

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å²)	Q<0.9
1	A	431/450 (95%)	0.36	15 (3%)	44	50	16, 23, 35, 41	0
1	B	425/450 (94%)	0.61	26 (6%)	22	25	17, 30, 42, 47	0
1	C	420/450 (93%)	0.39	14 (3%)	47	53	17, 27, 38, 44	0
1	D	427/450 (94%)	0.30	11 (2%)	56	62	16, 24, 36, 43	0
All	All	1703/1800 (94%)	0.41	66 (3%)	40	46	16, 26, 39, 47	0

The worst 5 of 66 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	125[A]	TYR	5.2
1	A	123	PRO	5.0
1	B	446	GLY	4.5
1	C	124	ILE	4.2
1	D	505	ASP	3.6

6.2 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors

of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors($Å^2$)	Q<0.9
2	UGA	A	6000	37/37	0.86	0.16	1.15	30,36,44,46	0
2	UGA	D	4000	37/37	0.97	0.09	-0.54	16,19,24,27	0
2	UGA	A	1000	37/37	0.97	0.10	-0.77	12,18,22,25	0
2	UGA	D	5000	37/37	0.92	0.11	-0.77	25,29,35,39	0
2	UGA	B	2000	37/37	0.96	0.09	-1.17	17,22,26,28	0
2	UGA	C	3000	37/37	0.97	0.09	-1.23	17,21,24,26	0

6.5 Other polymers (i)

There are no such residues in this entry.